EPoSS.
European Association on
Smart Systems Integration

Joint workshop by EPoSS and
INSIDE Industry Associations

# The Future of Innovation in Edge AI

Online, April 04, 2025

***Notes from the Workshop***

*Part VII: MultiSpin.AI: Overcoming Binary Barriers in Edge AI with Spintronics*

*Presented by Coordinator: Prof. Lior Klein.*

*Funded by the European Union under the EIC Pathfinder Open (Grant No. 101130046)*

## 1. Motivation: AI at the Edge Under Pressure

Artificial intelligence has become deeply integrated across domains – from natural language processing (e.g., ChatGPT, DeepSeek) to healthcare, education, automotive, and security systems. However, with this pervasive adoption comes a set of technical bottlenecks:

- **The end of Moore's Law**: The number of transistors in integrated circuits is no longer doubling reliably every two years.
- **The breakdown of Dennard Scaling**: As transistors shrink, their power density increases, escalating energy consumption.
- **The Von Neumann bottleneck**: The performance gap between CPU execution and memory access times limits throughput.

These challenges become **even more acute at the edge**, where AI must run in real time on constrained devices such as autonomous vehicles, drones, and wearable health monitors.

## 2. A New Computational Paradigm: Neuromorphic Computing

Neuromorphic computing mimics how the brain processes information, aiming to move beyond traditional digital computing architectures. The core of AI workloads involves **matrix-vector multiplications**, especially in neural networks. The standard approach – performing these calculations digitally – hits limits in speed and energy efficiency.

To address this, researchers have turned to **analog in-memory computing**, using **crossbars** to perform **multiplication and accumulation (MAC)** operations directly within memory arrays. This analog approach enables fast and energy-efficient processing, making it ideal for edge AI.

## 3. Crossbar Computing and Spintronic Limitations

A crossbar array uses a grid of vertical and horizontal lines with programmable resistors at each junction. These junctions store weights, and when voltage is applied, the resulting current reflects matrix-vector multiplication.

However, the performance of a crossbar is tightly linked to the type of resistive element used. One promising candidate is the **magnetoresistive tunnel junction (MTJ)**, used in **spintronics**.

Limitations of Conventional MTJs:

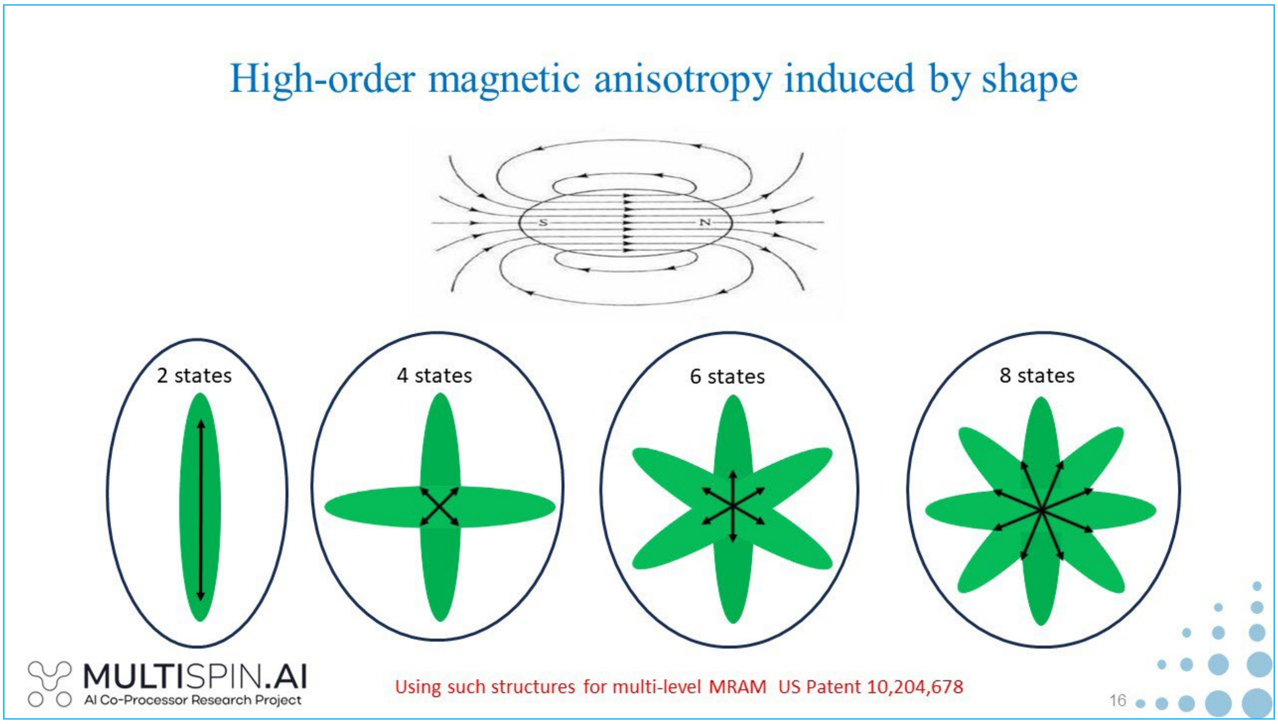
- MTJs have **only two resistance states** (parallel and anti-parallel magnetization).
- This **binary nature** restricts the precision of AI models.
- A study by Samsung using a 64×64 MTJ crossbar showed limited accuracy – especially for more complex classification tasks.

## 4. MultiSpin.AI: Introducing M²TJ for Multistate Magnetoresistance

MultiSpin.AI aims to revolutionize spintronic-based computing by replacing binary MTJs with **M²TJs** – devices that support **multiple discrete magnetic states**.

**Key Innovation:**

Instead of MTJs with a single axis of magnetization MultiSpin.AI introduces M2TJs for which there are multiple easy axes of magnetization at different angles within the same junction.






Multiple easy axes of magnetization are achieved by:

- **Shape anisotropy engineering**: Structuring ferromagnetic layers as **multiple intersecting ellipses**, each contributing a magnetic easy axis.
  - M2TJs are manipulated with spin currents which enables the stabilization of the following numbers of distinct magnetic configurations
  - **$2^4$ = 16** states (2 ellipses)
  - **$2^6$ = 64** states (3 ellipses)
  - **$2^8$ = 256** states (4 ellipses)

Although the number of **distinct magnetic configurations** exceeds the number of usable **resistance states**, it dramatically expands the representational capacity of a single junction.

## 5. From Concept to Fabrication: The Project Strategy

The MultiSpin.AI approach proceeds along three parallel development tracks:

**1. Single Layer Prototypes**

- Aim: Achieve over 16 distinct magnetic configurations.
- Use: Validate material and design feasibility.

**2. M²TJ Devices**

- Goal: Fabricate working magnetic tunnel junctions with at least 8 resistance states (equivalent to 3 bits per device).
- Importance: Maintain **discrete states** for improved accuracy and reproducibility – unlike continuous-state devices prone to drift.

**3. Crossbar Arrays**

- Objective: Build a functional **4×4 M²TJ crossbar** where each M2TJ supports 4 resistance states to demonstrate integration potential.
- Use Case: Run actual AI tasks to validate performance, energy efficiency, and accuracy.

In parallel, the team is developing custom **algorithms and a PLC-based control framework** to interface the hardware with AI workloads.

**6. Consortium and Roles**

- MultiSpin.AI unites top research institutes and innovators in magnetism, microelectronics, and AI:**Bar-Ilan University (Israel)** – Project coordination, modelling, fabrication and characterization
- **INESC-MN (Portugal)** – Main fabrication effort
- **UC Louvain (Belgium)** – Algorithms and characterization
- **SpinEdge (Israel)** – Algorithms and prototype development
- **I-FEVS (Italy)** – Prototype development and testing.
- **Amires (Czech Republic )** – Dissemination and exploitation management

## 7. Why It Matters: Overcoming Bottlenecks with MultiState Logic

The binary nature of current MTJ-based crossbars limits their precision and scalability. MultiSpin.AI directly addresses this:

- More resistance states = **greater model accuracy**
- Improved accuracy = **smaller required crossbar size = lower power consumption**
- Discrete multistate behavior = **higher stochastic reliability** (compared to continuous-state memory devices)

If successful, the project could significantly **increase energy efficiency**, with the potential to exceed **hundreds of tera-operations per watt**.

To better understand the impact of MultiSpin.AI, the following table presents a comparison with similar approaches from recent research:

**Multispin.ai Performance Comparison with Other Neuromorphic Technologies**

| Technology | Energy Efficiency (TOPS/W) | Comment |
|---|---|---|
| **MultiSpin.AI (2024)** | >1,000 TOPS/W | High-speed, ultra-efficient spintronic crossbar AI processing of mainstream AI models. |
| **Gaussian Probabilistic Bits - Stochastic Spintronics (Phys.org, 2024)**[1] | ~100 TOPS/W | Probabilistic AI models with stochastic spintronics aiming NP problems |

1 Tohoku University (Japan) and the University of California, Santa Barbara (UCSB, USA). Phys.org (2024) Energy-Efficient Probabilistic Computing Using Gaussian Probabilistic Bits from Stochastic Spintronics Devices.

| Technology | Energy Efficiency (TOPS/W) | Comment |
|---|---|---|
| **Neuromorphic Computing with Spintronics (Nature Reviews Materials, 2023)**[2] | 300-500 TOPS/W | Improved memory access and hybrid computing strategies |
| **IBM TrueNorth (2014)**[3] | 600+ TOPS/W | First large-scale neuromorphic chip with event-driven processing; however, it suffers from lack of on-chip training, reliance on spiking neural networks (SNNs), and low computational precision (1-bit synapses), limiting its adaptability and mainstream AI compatibility |
| **Intel Loihi-2 (2017)**[4] | ~15 TOPS/W | Low-power AI processing reliance on spiking neural networks (SNNs), does not natively support mainstream AI models |

## 8. Challenges and the Path Forward

The MultiSpin.AI project is poised to redefine the landscape of AI hardware by providing a scalable **energy-efficient alternative** to existing digital accelerators. The anticipated benefits include:

- **Drastic energy savings**, reducing the environmental footprint of AI computations.
- **Faster AI inference times**, unlocking new possibilities for real-time applications in robotics, healthcare, and edge intelligence.
- **Enhanced scalability**, allowing AI models to process vast amounts of data without the limitation imposed by traditional von Neumann architectures.

By merging **spintronics with AI**, MultiSpin.AI sets the stage for **next-generation computational paradigms** that are not only more sustainable but also more powerful and versatile than current solutions. As the project advances towards commercialization, it promises to bring a **transformation shift in AI processing**, enabling the next wave of intelligent, power-efficient, and high-performance computing systems.

While promising, the fabrication of **M$^2$TJ devices** is technically complex – especially within the constraints of university labs. To fully scale and deploy the technology, partnerships with **European chip manufacturers** will be key.

This collaboration would:

- Accelerate fabrication
- Enable larger-scale integration
- Position Europe as a leader in **energy-efficient AI hardware**

2 Nature Reviews Materials (2023) Neuromorphic Computing with Spintronics: State-of-the-Art and Future Perspectives.

3 IBM Research (2014) TrueNorth: A Neuromorphic CMOS Chip for AI Processing.

4 Intel Labs (2017) Loihi: A Neuromorphic Computing Chip

## Conclusion: MultiSpin.AI’s Vision for Edge AI

MultiSpin.AI offers a new route to break through the limitations of conventional AI hardware. By integrating multistate spintronics into crossbar architectures, the project promises:

- Improved **accuracy**
- Enhanced **energy efficiency**
- Reduced **hardware footprint**

As AI continues to push toward the edge, these advances could be transformative.

To follow project updates, visit: www.multispinai.eu